

The MPEG-7 Visual Standard for Content Description—An Overview

Thomas Sikora, *Senior Member, IEEE*

Abstract—The MPEG-7 Visual Standard under development specifies content-based descriptors that allow users or agents (or search engines) to measure similarity in images or video based on visual criteria, and can be used to efficiently identify, filter, or browse images or video based on visual content. More specifically, MPEG-7 specifies color, texture, object shape, global motion, or object motion features for this purpose. This paper outlines the aim, methodologies, and broad details of the MPEG-7 Standard development for visual content description.

Index Terms—Coding, descriptors, MPEG-7, similarity-based retrieval, standardization, visual information.

I. INTRODUCTION

RECENT years have seen a rapid increase in volume of image and video collections. A huge amount of information is available, and every day, gigabytes of new visual information is being generated, stored, and transmitted. However, it is difficult to access this visual information unless it is organized in a suitable way—to allow efficient browsing, searching, and retrieval. Image retrieval has been a very active research and development domain since the early 1970s. During the early 1990s—with the advent of digital video—research on video retrieval became of equal importance. A very popular means for image or video retrieval is to annotate images or video with text, and to use text-based database management systems to perform image retrieval. However, text-based annotation has significant drawbacks when confronted with large volumes of images. Annotation can in these circumstance become significantly labor intensive. Furthermore, since images are rich in content, text may in many applications not be rich enough to describe images. To overcome these difficulties in the early 1990s, content-based image retrieval emerged as a promising means for describing and retrieving images. Content-based image retrieval systems describe images by their own visual content rather than text, such as color, texture, and objects' shape information.

In the late 1990s—with the large scale introduction of digital images and video to the market [1]—the necessity for interworking between image/video retrieval systems of different vendors arose. For this purpose, in 1997 the ISO MPEG Group initiated the “MPEG-7 Multimedia Description Language” work item. The target of this activity was to issue an international MPEG-7 Standard, defining standardized descriptions and description systems that allow users or agents

to search, identify, filter, and browse audiovisual content [2], [3]. MPEG-7 is currently still under definition and will become international standard in July 2001. Besides support for meta-data and text descriptions of the audiovisual content, much focus in the development on MPEG-7 has been in the definition of efficient content-based description and retrieval specifications.

The purpose of this paper is to provide a broad overview of the MPEG-7 content-based visual descriptors. For an overall overview of the MPEG-7 Standard, more detailed descriptions of MPEG-7 content-based visual, audio and speech descriptors, the reader is referred to the literature in [2], [3] and to the remaining papers of this Special Issue on MPEG-7 in [5]–[12].

II. SCOPE OF MPEG-7 VISUAL STANDARD

The ultimate goal and objective of MPEG-7 Visual Standard is to provide standardized descriptions of streamed or stored images or video—standardized header bits (visual low-level *Descriptors*) that help users or applications to identify, categorize or filter images or video. These low-level descriptors can be used to compare, filter, or browse image or video purely based on nontext visual descriptions of the content, or if required, in combination with common text-based queries. Because of their descriptive features, the challenge for developing such MPEG-7 Visual nontext descriptors is that they must be meaningful in the context of various applications. They will be used differently for different user domains and different application environments.

Selected application examples include digital libraries (image and video catalogue), broadcast media selection (TV channels), and multimedia editing (personalised electronic news service, media authoring). Among this diversity of possible applications, the MPEG-7 Visual feature descriptors allow users or agents to perform the following tasks taken as examples.

- 1) *Graphics*: Draw a few lines on a screen and get, in return, a set of images containing similar graphics or logos.
- 2) *Images*: Define objects, including color patches or textures, and get, in return, examples among which you select the ones of interest.
- 3) *Video*: On a given set of video objects, describe object movements, camera motion, or relations between objects and get, in return, a list of videos with similar or dissimilar temporal and spatial relations.
- 4) *Video Activity*: On a given video content, describe actions and get a list of videos where similar actions happen.

The MPEG-7 Visual *Descriptors* describe basic audiovisual content of media based on visual information. For images and video, the content may be described, for example by the shape

Manuscript received January 2, 2001; revised March 12, 2001.

The author is with the Heinrich-Hertz-Institute for Communication Technology (HHI), Interactive Media—Human Factors, D-10587 Berlin, Germany (e-mail: sikora@hhi.de).

Publisher Item Identifier S 1051-8215(01)04986-2.

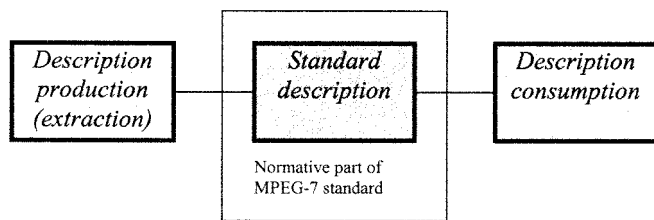


Fig. 1. Scope of MPEG-7.

of objects, object size, texture, color, movement of objects, and camera motion. These descriptors need to be extracted from the image or video content. It should be noted that MPEG-7 descriptor data may be physically located with the associated AV material, in the same data stream, or on the same storage system. Alternatively, the descriptors could also be located anywhere else, as long as it is possible to link AV material and their MPEG-7 descriptions efficiently. Once the MPEG-7 descriptors are available, suitable search engines can be employed to search, filter, or browse visual material based on suitable similarity measures. It must be noted that practical search engine implementations may need to match content based on a weighted combination of visual descriptors (i.e., color and texture) and maybe also including common text-based queries. For this purpose, MPEG-7 defines *Description Schemes* that describe combinations of descriptors in a suitable way. The reader is referred to references in [2], [3] to obtain a broad overview of the various parts of the MPEG-7 Standard and what role the MPEG-7 Visual descriptors take within the overall MPEG-7 scenario.

Fig. 1 depicts the MPEG-7 processing chain to explain the scope of the MPEG-7 Standard in a simple way. A typical application scenario involves that MPEG-7 descriptors are extracted (produced) from the content. It is important to understand that for most visual descriptors, the MPEG-7 Standard only partly describes how to extract these features, to leave a maximum of flexibility to the various applications. Extraction for most parts of the MPEG-7 Visual Standard is thus not normative. However, as a normative requirement, the semantics and the syntax of the descriptors must conform to the MPEG-7 specification in order to be compliant to the MPEG-7 Standard. Compliant MPEG-7 descriptors can be accessed, understood, and consumed by applications that are able to decode and process MPEG-7 features. How the MPEG-7 descriptors are being used for further processing—i.e., for search and filtering of content—is again not specified by MPEG-7, to leave maximum of flexibility to applications. In particular, how similarity between images or video is defined is left to the specific applications requirements.

III. DEVELOPMENT OF THE STANDARD

MPEG develops specifications based on a well defined standard development framework. When a standard activity such as MPEG-7 is agreed within MPEG, the following steps are followed: requirements for technology is specified, technology is requested through an official “Call for Proposals,” and the technology proposed to MPEG is evaluated by MPEG experts based on pre-defined performance criteria. After the best technology has been identified, the MPEG Standard is being developed in

a collaborative effort through the definition of an Experimentation Model and a series of Core Experiments. This procedure already proved successful in the course of the development of the MPEG-1, MPEG-2, and MPEG-4 Standards [1], [4].

The purpose of an Experimentation Model (XM) within MPEG-7 is to specify and implement the feature extraction, encoding and decoding algorithms, as well as search engines. This supports execution of “Core Experiments” under controlled conditions in a common environment. Collaborative experiments performed by multiple independent parties can thus produce identical results. An XM specifies the input and output formats for the uncoded data, the extraction method used to obtain a descriptor, and the format of the bitstream containing the data. It also specifies the algorithm for decoding and defines suitable metrics for comparing content as described by descriptors of same type. The XM exists, and will serve this function, even after the standard is defined.

A Core Experiment aims to improve to the current technology in the XM. It is defined with respect to the XM, which includes the Common Core algorithm. A Core Experiment proposal describes a potential algorithmic improvement to the XM, i.e., a texture descriptor different to the one defined by the XM. Furthermore, the full description of extraction and search algorithms, and the specification of experimental conditions (such as image data set, similarity measures, etc.) to compare the proposed Core Experiment technique against the performance of the XM, are provided. A Core Experiment is established by the MPEG Video group if two independent parties are committed to perform the experiment. If a Core Experiment is successful in improving on a technique described in the XM in terms of retrieval efficiency, provisions for functionalities not supported by the XM, and implementation complexity, with the successful technique, is incorporated into the newest version of the XM. The technique may either replace an existing technique, or supplement the algorithms already supported by the XM. Core Experiments are being performed between two MPEG Video group meetings. At each MPEG Video group meeting, the results of the Core Experiments are reviewed, the XM is updated accordingly, and a new version of the XM is released.

Various proposals for color, texture, shape/contour, and motion descriptors were evaluated in performance tests in 1998, and the most promising proposals were adopted for the first visual XM. In subsequent meetings, these first tools were improved in the Core Experiment process by introduction of refinements and new promising algorithms. This refinement process took until early 2001, when the successful descriptors were finally adopted for the MPEG-7 Visual Standard.

IV. VISUAL DESCRIPTORS FOR IMAGES AND VIDEO

The MPEG-7 descriptors that were developed can be broadly classified into general visual descriptors and domain specific visual descriptors. The former include color, texture, shape, and motion features, while the latter are application dependent and include identification of human faces and face recognition. Since the standardization of the domain-specific descriptors is still under development, this paper concentrates on those general descriptors which can be used in most applications.

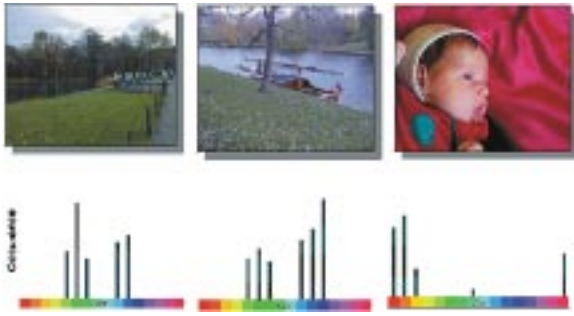


Fig. 2. Three color images and their MPEG-7 histogram color distribution, depicted using a simplified color histogram. Based on the color distribution, the two left images would be recognized as more similar compared to the one on the right.

A. Visual Color Descriptors

Color is one of the most widely used visual features in image and video retrieval. Color features are relatively robust to changes in the background colors and are independent of image size and orientation. Color descriptors can be used for describing content in still images and video, respectively.

Considerable design and experimental work, and rigorous testing, have been performed in MPEG-7 to arrive at efficient color descriptors for similarity matching. No single generic color descriptor exists that can be used for all foreseen applications. As a result, a range of descriptors has been standardized, each suitable for achieving specific similarity-matching functionalities. In the following, a brief overview of each descriptor is provided. For more details, the reader is referred to [5], [6].

Color Spaces: To allow interoperability between various color descriptors, normative color spaces are constrained to hue-saturation-value (HSV) and hue-min-max-diff (HMMD). HSV is a well-known color space widely used in image applications. HMMD is a new color space defined by MPEG and is only used in the color structure descriptor (CSD) explained below.

Scalable Color Descriptor (SCD): One of the most basic description of color features is provided by describing color distribution in images. If such a distribution is measured over an entire image, global color features can be described. Fig. 2 depicts examples of color images and their respective color distributions in a color histogram. The MPEG-7 generic SCD is a color histogram encoded by a Haar transform. It uses the HSV colors space uniformly quantized to 255 bins. To arrive at a compact representation the histogram bin values are nonuniformly quantized in a range from 16 bits/histogram for a rough representation of color distribution and up to 1000 bits/histogram for high-quality applications. Matching between SCD realizations can, i.e., be performed by matching Haar coefficients or histogram bin values employing an L1 norm.

Dominant Color Descriptor: This color descriptor aims to describe global as well as local spatial color distribution in images for high-speed retrieval and browsing. In contrast to the Color Histogram approach, this descriptor arrives at a much more compact representation—at the expense of lower performance in some applications. Colors in a given region are clustered into a small number of representative colors. The descriptor consists of the representative colors, their percentages in a region, spatial coherency of the color, and color variance.

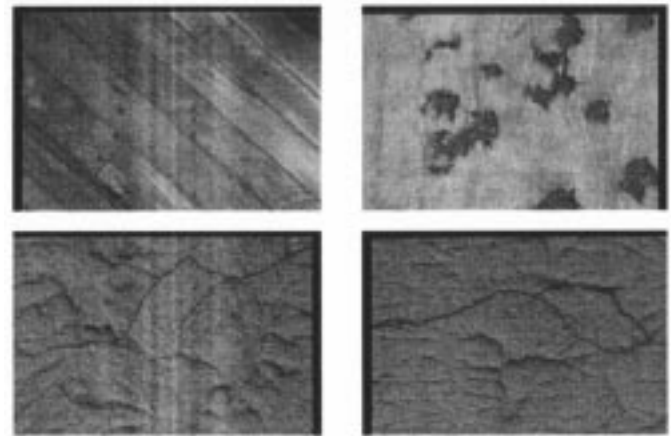


Fig. 3. Examples of grayscale images with different textures. Using the MPEG-7 Visual texture descriptors, the two images on the bottom would be rated of similar texture, while less similar in texture compared to the two images on the top.

Color Layout Descriptor: This descriptor is designed to describe spatial distribution of color in an arbitrarily-shaped region. Color distribution in each region can be described using the Dominant Color Descriptor above. The spatial distribution of color is an effective description for sketch-based retrieval, content filtering using image indexing, and visualization.

CSD: The main purpose of the CSD is to express local color features in images. To this aim, a 8×8 pel structuring block scans the image in a sliding window approach. With each shift of the structuring element, the number of times a particular color is contained in the structure element is counted, and a color histogram is constructed in such a way.

Group-of-Frames/Group-of-Pictures (GoF/GoP) Color Descriptor: The GoF/GoP color descriptor defines a structure required for representing color features of a collection of similar frames or video frames by means of the SCD. It is useful for retrieval in image and video databases, video shot grouping, image-to-segment matching, and similar applications. It consists of average, median, and intersection histograms of groups of frames calculated based on the individual frame histograms.

B. Visual Texture Descriptors

Texture refers to the visual patterns that have properties of homogeneity or not, that result from the presence of multiple colors or intensities in the image. It is a property of virtually any surface, including clouds, trees, bricks, hair, and fabric. It contains important structural information of surfaces and their relationship to the surrounding environment. Describing textures in images by appropriate texture descriptors provides powerful means for similarity matching and retrieval. To illustrate texture properties a collection of images with different textures is depicted in Fig. 3. MPEG-7 has defined appropriate texture descriptors that can be employed for a variety of applications and tasks [5], [6]. For similarity retrieval texture descriptors can be used independent or in combination with other MPEG-7 descriptors.

Homogenous Texture Descriptor: The Homogenous Texture Descriptor describes directionality, coarseness, and regularity of

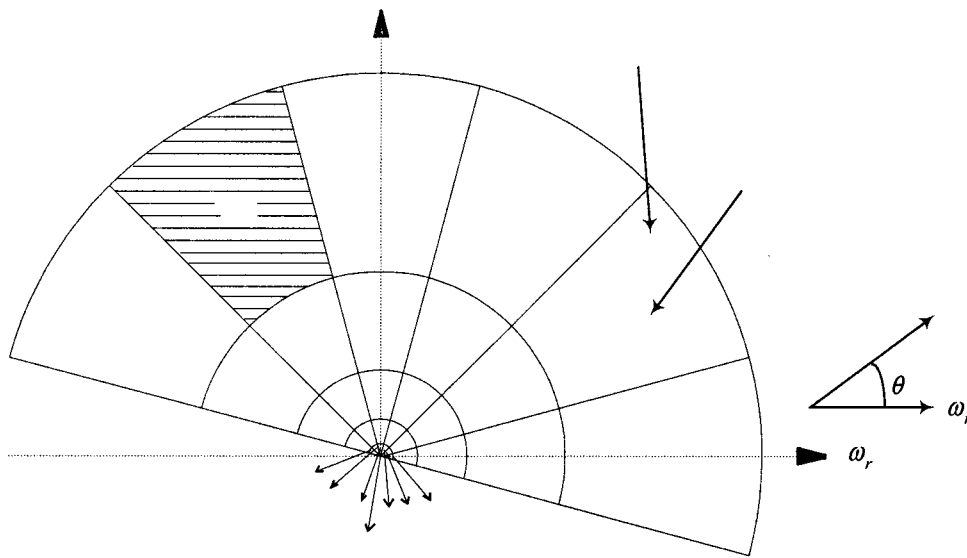


Fig. 4. Frequency layout for MPEG-7 Homogenous Texture Descriptor frequency extraction. Energy and energy deviation values are extracted from this frequency division into 30 channels.

patterns in images and is most suitable for a quantitative characterization of texture that has homogenous properties. It can be used for similarity image-to-image matching for texture image databases. In order to describe the image texture, energy, and energy deviation, values are extracted from a frequency layout. The descriptor is based on a filter bank approach employing scale and orientation sensitive filters. Suitable descriptions are obtained in the frequency domain by computing mean and standard variation of frequency coefficients. A radon transform followed by Fourier transform can be employed to achieve suitable computational efficiency for low complexity applications. To arrive at scale and rotation-invariant description and matching of texture, the frequency space is partitioned into 30 channels with equal division in the angular direction and octave division in radial direction (see Fig. 4). The individual feature channels are filtered using 2-D-Gabor functions.

Non-Homogenous Texture Descriptor (Edge Histogram): In order to also provide descriptions for nonhomogenous texture images, MPEG-7 defined an Edge Histogram Descriptor. This descriptor captures spatial distribution of edges, somewhat in the same spirit as the Color Layout Descriptor. The extraction of this descriptor involves division of image into 16 nonoverlapping blocks of equal size. Edge information is then calculated for each block in five edge categories: vertical, horizontal, 45° , 135° , and nondirectional edge. It is expressed as a 5-bin histogram, one for each image block. The descriptor is scale invariant, and supports both rotation-sensitive and rotation-invariant matching. It is also very compact because each histogram bin is nonuniformly quantized using 3 bits, resulting in a descriptor size of 240 bits.

C. Visual Shape Descriptors

In many image data-base applications, the shape of image objects provides a powerful visual clue for similarity matching. Typical examples of such applications include binary images with written characters, trademarks, pre-segmented object contours and 2-D/3-D virtual object boundaries. In image retrieval,

it is usually required that the shape descriptor is invariant to scaling, rotation, and translation. Shape information can be 2-D or 3-D in nature, depending on the application. In general, 2-D shape description can be divided into two categories, contour-based and region-based. The former uses only boundary information of objects, suitable to describe objects that have similar contour characteristics. The latter uses the entire shape region to extract a meaningful description which is most useful when objects have similar spatial distributions of pixels in objects. Dependent on the application or objects characteristics, it is useful to employ either region- or contour-based descriptors. In the following the MPEG-7 Visual shape descriptors are briefly introduced. For more details, the reader is referred to [5], [7].

3-D Shape Descriptor—Shape Spectrum: The MPEG-7 3-D shape descriptor is useful to compare natural or virtual 3-D objects. The descriptor is based on a shape spectrum concept. Roughly speaking, the shape spectrum is defined as the histogram of a shape index, computed over the entire 3-D surface. The shape index itself measures local convexity of each local 3-D surface. Histograms with 100 bins are used—each quantized by 12 bits.

Region-Based Descriptor—Art: The MPEG-7 Region-Based Descriptor ART (Angular Radial Transformation) belongs to the class of moment invariants methods for shape description. This descriptor is suitable for shapes that can be best described by shape regions rather than contours. The main idea behind moment invariants is to use region-based moments which are invariant to transformations, as the shape feature. The MPEG-7 ART descriptor employs a complex Angular Radial Transformation defined on a unit disk in polar coordinates to achieve this goal. Coefficients of ART basis functions are quantized and used for matching. The descriptor is very compact (140 bits/region) and also very robust to segmentation noise. Examples of similarity matching between various shapes using the ART descriptor is shown in Fig. 5.

Contour-Based Shape Descriptor: Objects for which shape features are best expressed by contour information can be de-

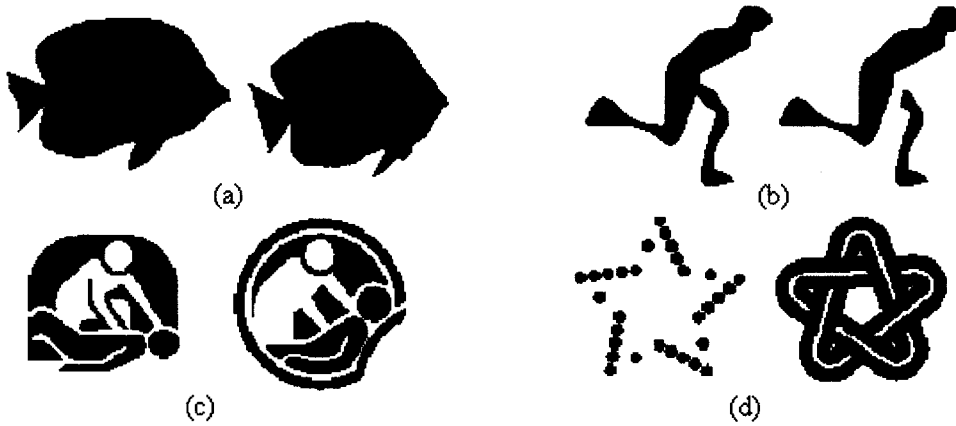


Fig. 5. Examples of various shapes that can be indexed using MPEG-7 Region-Based Shape Descriptor. Images contained in either of the sets (a)–(d) would be rated similar and dissimilar to the ones in the remaining sets. For example, images in set (a) would be identified being similar and dissimilar to the ones in set (b), (c), or (d).

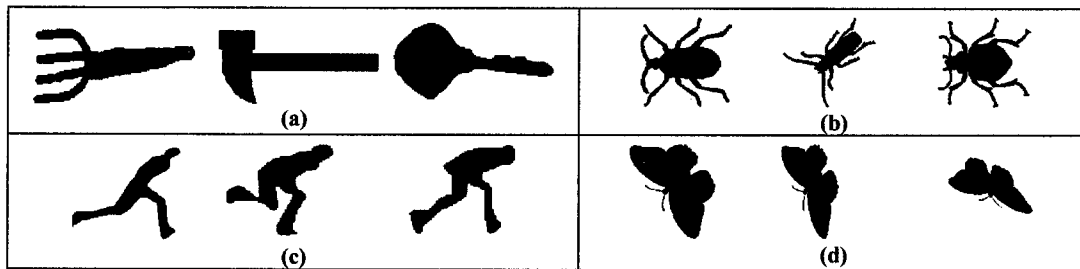


Fig. 6. Examples of shapes that can be indexed using MPEG-7 Contour-Based Shape Descriptor.

scribed using the MPEG-7 Contour-Based Descriptor. This descriptor is based on curvature scale-space (CCS) representations of contours and also includes of eccentricity and circularity values of the original and filtered contours. A CCS index is used for matching and indicates the heights of the most prominent peak, and the horizontal and vertical positions on the remaining peaks in the so-called CSS image. The average size of the descriptor is 122 bits/contour. Fig. 6(b)–(d) depicts similarity matching results using the MPEG-7 Contour-Based Shape Descriptor. Fig. 6(a) shows examples of shapes which have similar region but different contour properties. Such objects would be considered as very different by the contour-based shape descriptor.

2-D/3-D Shape Descriptor: In general, the shape of a 3-D object can be described approximately by a limited number of 2-D shapes which are taken as 2-D snapshots from different viewing angles. Thus the MPEG-7 2-D shape descriptors can be used to describe each of the 2-D shapes taken as snapshots from the 3-D object. A similarity matching between 3-D objects thus involves matching multiple pairs of 2-D views taken one from each of the objects. In general, good performance for 3-D shapes has been demonstrated using the MPEG-7 2-D Contour-Based Descriptor.

D. Motion Descriptors for Video

All MPEG-7 descriptors described above for color, texture, and shape of objects can be readily employed to index images in video sequences. As an example, color and texture descriptors are frequently used for temporal video segmentation, such as

shot boundary detection, localization of camera dissolve effects, etc. Description of motion features in video sequences can provide even more powerful clues regarding its content. In general, describing motion in video sequences by motion fields can be very expensive in terms of bits per image, even if motion vector fields are coarse. MPEG-7 has developed descriptors that capture essential motion characteristics from the motion field into concise and effective descriptions. The most dominant characteristics are provided by camera motion and object motion descriptors [5], [8].

Motion Activity Descriptor: For a video segment (i.e., a given group of frames) its overall activity level or pace of motion or action in a scene is captured by the Motion Activity Descriptor. This descriptor describes whether a scene is likely to be perceived by a viewer as being slow, fast paced, or action paced. Examples of high “activity” includes scenes such as soccer or baseball matches. On the other hand scenes such as taken from a TV news reader would be low-level action shots. The motion activity intensity descriptor enables selection of video segments or of a program that can be categorized by such attributes. The descriptor measures the intensity of motion based on standard deviations of motion vector magnitudes. The standard deviations are quantized into five activity values. Optionally, the direction of motion, spatial distribution of motion activity, and the temporal distribution of motion activity can also be extracted as descriptions and used for similarity matching.

Camera Motion Descriptor: The movement of a camera or of the virtual viewpoint in a scene is described by the Camera Motion Descriptor (Fig. 7). This descriptor details what kind of

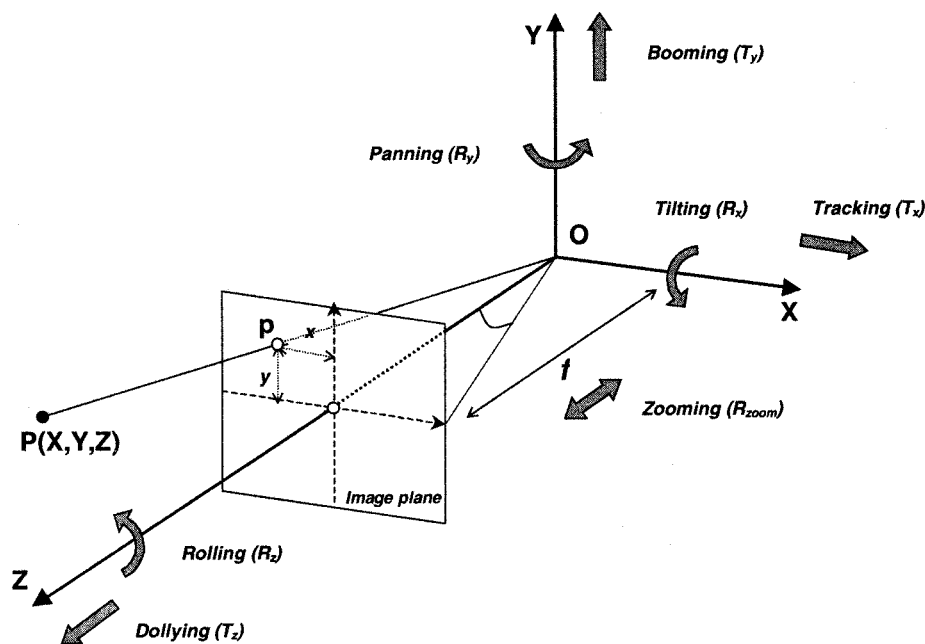


Fig. 7. Camera model for MPEG-7 Camera Motion Descriptor. Perspective projection to image plane p and camera motion parameters. The (virtual) camera is located in O .

global motion parameters are present at what instance in time in a scene. Such parameters may, in some applications, be provided directly by the camera. It is also possible to estimate such camera parameters from the image pixel intensities using an appropriate camera model and motion-estimation techniques. The descriptor can be used to query video sequences based on certain global motion parameters, such as, i.e., scenes with a lot of zoom activity or scenes with mainly translatory motion. Also, queries that allow the similarity matching of motion in particular time periods is possible (i.e., whether translatory camera motion occurred between the 10th–20th second in a video sequence).

Warping Parameters: Another way to evaluate motion in a scene is based on a Warping Parameter Descriptor. Here, the global motion is expressed with reference to a global Sprite or Mosaic of a video scene [4].

Motion Trajectory Descriptor: Motion can also be described for each independently moving object in a video sequence using the Motion Trajectory Descriptor. This descriptor details the displacement of objects over time as described by an object motion trajectory. It allows matching of object trajectories for motion-based video querying, video hyper-linking, and related applications. For example, in video-based traffic surveillance applications, the motion of cars and pedestrians can be described and stored using MPEG-7 Motion Trajectory Descriptor. Queries can be performed based on the descriptors to allow a fast access to scenes with required object motion. Possible queries include search for “objects that pass near a specific area” or “objects moving faster than the speed limit.”

V. SUMMARY AND CONCLUSION

In this paper, the aim, development methodologies, and some details of the MPEG-7 Standard for visual content description was outlined. The MPEG-7 Visual standard specifies content-

based descriptors that can be used to efficiently identify, filter, or browse images or video based on visual content rather than text. MPEG-7 descriptors are extracted from images or video sequences using suitable extraction methods and can be stored or transmitted entirely separate from the media content. The descriptors allow users or agents (or search engines) to evaluate similarity in images or video based on color, texture, object shape, global motion, or object motion features. The implementation of descriptors compliant to MPEG-7 specifications is normative for applications which intent to provide interoperability with other MPEG-7 implementations. However, to allow a maximum of flexibility for the design of applications, MPEG-7 specifies only minor parts of the descriptor extraction methods, and not how similarity between content should be measured.

The visual MPEG-7 descriptors detailed in this paper can be readily employed on their own to built powerful applications. It is worthwhile, however, to also view the visual descriptors in the context of the remaining MPEG-7 Standards developments [2], [3]. On the lowest abstraction level, the audiovisual *descriptors* describe basic audiovisual content of media. Similar to video, the audio or speech content may be described by key, mood, tempo, tempo changes, and position in sound space [9]. A limited amount of pre-defined *Description Schemes* provide semantic, as well as nonsemantic, information about audio, speech, image, or video content for specific applications [10]. Such description schemes also allow to make queries on combinations of low-level visual descriptors, provide for text annotation of media and for metadata description of content. The MPEG-7 *Description Definition Language* allows to flexibly define MPEG-7 Description Schemes based on XML schema [11]. Finally, how descriptors or description schemes are being stored, accessed or transmitted will be specified in the MPEG-7 Systems specification [12].

ACKNOWLEDGMENT

The author is gratefully indebted to the many individuals and companies who have been instrumental in defining and verifying the details of the MPEG-7 Visual Standard. He would like to thank Dr. M. Bober for his very valuable help with the manuscript, and the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

- [1] R. Schäfer and T. Sikora, "Digital video coding standards and their role in video communications," *Proc. IEEE*, vol. 83, pp. 907–924, June 1995.
- [2] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Standard*. New York: Wiley, 2001.
- [3] S.-F. Chang, T. Sikora, and A. Puri, "Overview of MPEG-7 Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 688–695, June 2001.
- [4] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 19–31, Feb. 1997.
- [5] *MPEG-7 Visual Part of Experimentation Model Version 8.0*, A. Yamada, M. Pickering, S. Jeannin, L. Cieplinski, J.-R. Ohm, and M. Editors, Eds., Oct. 2000, ISO/IEC JTC1/SC29/WG11/N3673.
- [6] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "MPEG-7 color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 703–715, June 2001.
- [7] M. Bober, "MPEG-7 Visual Shape Descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. xref–x, June 2001.
- [8] S. Jeannin and A. Divakaran, "MPEG-7 visual motion descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 720–724, June 2001.
- [9] S. Quackenbush and A. Lindsay, "Overview of MPEG-7 audio," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 725–729, June 2001.
- [10] P. Salembier and J.-R. Smith, "MPEG-7 multimedia description schemes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 748–759, June 2001.

[11] J. Hunter, "An overview of the MPEG-7 description definition language (DDL)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 765–772, June 2001.

[12] O. Avaro and P. Salembier, "MPEG-7 Systems: Overview," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 760–764, June 2001.



Thomas Sikora (M'93–SM'96) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from Bremen University, Germany, in 1985 and 1989, respectively.

In 1990, he joined Siemens Ltd., Melbourne, Australia, and Monash University, Melbourne, Australia, as a Project Leader, responsible for video compression research activities in the Australian "Universal Broadband Video Codec" consortium. Between 1994 and 1998, he directed video-coding research activities at Heinrich-Hertz-Institute (HHI), Berlin, Germany. He is currently Head of the Interactive Media—Human Factors Department at HHI, in addition to being an Adjunct Professor in the Electrical Engineering Department, Technical University Berlin, Berlin, Germany. He is co-founder of 2SK Media Technologies and Perspective Technologies GmbH, two recent Berlin-based start-up companies involved in audio and video signal processing, compression technology and 3-D visualization. He has been involved in ITU and ISO standardization activities, as well as in several European research activities for a number of years. As the chairman of the ISO-MPEG video group, he is responsible for the development and standardization of the MPEG video-coding algorithms. Between 1995–1998, he also served as the chairman of the European COST 211^{tr} video compression research group. He is frequently an industry consultant on issues related to interactive digital audio and video.

Dr. Sikora was the recipient of the 1996 German ITG Award (German Society for Information Technology). He is an Associate Editor for a number of international journals, including the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *IEEE Signal Processing Magazine*, the *EURASIP Signal Processing Journal*, and the *EURASIP Signal Processing: Image Communication Journal*. He is a member of ITG.